DOCUMENT RESUME

ED 275 167 FL 016 071

AUTHOR Boeijen, Marijke
TITLE Testing Listening Comprehension: Notions and Functions as "Discrete Points" in a Communicative Context.
PUB DATE 84
NOTE 16p.; In: Practice and Problems in Language Testing. Papers from the International Symposium on Language Testing (7th, Colchester, England, 1984); see FL 016 066.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Communicative Competence (Languages); *Criterion Referenced Tests; Difficulty Level; Foreign Countries; French; *Language Proficiency; *Language Tests; *Listening Comprehension; Notional Functional Syllabi; *Test Construction; Test Use
IDENTIFIERS Netherlands; *Threshold Level

ABSTRACT

This study examines two kinds of French language tests originating from the Threshold-Level (T-Level) or Niveau-Seuil theory of language learning in the context of several linguistic and testing concepts: discrete-point testing with regard to notions and functions from T-Level theory, communicative context, and assessment of listening comprehension. The first test type is objective-based or criterion-referenced, having a formative purpose and intended for the first years of the language learning period. The learning objectives on which these tests are based are determined from general notions as described in T-Level theory, with the discrete-point approach being predominant, while the communicative aspect is limited. The second test type is proficiency-oriented with a summative function to be used at the end of the first language learning period of about 4 years. These tests cover parts of both the specific notions and language functions of the T-Level. The communicative aspect is more evident here, and although the tests themselves deal with discrete points of linguistic competence, they are embedded in more holistic comprehension procedures. Some scores resulting from the tests' use are included. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275167

TESTING LISTENING COMPREHENSION: NOTIONS AND FUNCTIONS AS "DISCRETE POINTS" IN A COMMUNICATIVE CONTEXT

FL 016 071

MARIJKE BOEIJEN
(NATIONAL INSTITUTE FOR EDUCATIONAL MEASUREMENT, CITO, ARNHEM)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. Culhane

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TESTING LISTENING COMPREHENSION: NOTIONS AND FUNCTIONS AS "DISCRETE POINTS" IN A COMMUNICATIVE CONTEXT

Marijke Boeijen (National Institute for Educational Measurement, Cito, Arnhem)

1 Introduction

In this paper two kinds of French tests will be presented that both originated from The Threshold-Level (T-Level) or Un Niveau-Seuil. We shall first have a short look at some of the concepts mentioned in the title of this paper; especially the concept of discrete point-testing with regard to Notions and Functions from the T-Level will be discussed. The concept of communicative context and the assessment of listening comprehension will be paid attention to. Specimens of the tests will be demonstrated and back-ground information will be given. Some psychometric data will also be presented.

The first type of tests are objective-based or criterion-referenced tests which have a formative purpose and are intended for the first years of the learning period. The learning objectives have been determined on the basis of the General Notions as described in the T-Level or Un Niveau-Seuil. The discrete-point approach is rather predominant here, whereas the communicative aspect is limited and could be characterized as pre-communicative.

The second type concerns proficiency-oriented tests with a summative function, at the end of the first period of circa 4 years. These tests cover parts of the Specific Notions on the one hand and of the Language-Functions on the other. The communicative aspect here is more evident and although the tests themselves deal with discrete points of the linguistic competence, these points are embedded in more holistic comprehension procedures.

2 Discrete-point testing and listening comprehension

When Cito started the construction of objective-based tests for listening comprehension the current tests in use for beginning students in foreign languages were of the discrete-point type: They usually measured isolated elements of the linguistic competence focusing on specific features of grammar. This procedure of breaking up the language into its smallest constituents cannot be used either in teaching or in testing listening comprehension. In a communicative skill as is listening comprehension, it is no use asking for the simple knowledge of constituting elements. Context in the largest sense of the word plays an indispensable part.

Still the need was felt to structure the complex listening-process into smaller units. It seemed possible to do so once the T-Level-model was published. In this model language-learning is defined in terms of practical communicative ability. The analysis made

on the basis also of extra-linguistic features (target-group, language activities, settings, roles, topics) led to an interesting categorization of the language-forms or exponents into General/ Specific Notions and Functions. (Van Ek 1976) Within the three main categories one can distinguish smaller units which can be seen as learning objectives. This kind of objectives seems appropriate to the teaching and testing of language in terms of skills. More particularly the General Notions (e.g. temporal, spatial, logical etc., see 5.1) constitute a finite domain. As far as comprehension is concerned, the frequent occurence of General Notions necessitates good mastery of their language-forms. Now these language-forms, centered round a common denominator that we will call a learning objective and which in fact is a subcategory of the General Notions, can be considered to be discrete points.

3 Notions and Functions in a communicative context

Measuring listening comprehension in a communicative context should not be equated with measuring communicative competence, although there is some relationship. The latter is a rather ambiguous concept open to different interpretations. (Canale, Swain 1980) Without going into a discussion on what aspects of human behaviour have to be included in its definition, we may safely say that communicative competence includes abilities which go beyond linguistic competence proper. However, any language-related communicative situation requires a certain level of linguistic competence. Even a if non-native speaker replaces his lack of command of the foreign language by gesticulation when he wants to express himself, he still has no control at all over the linguistic reaction of the respondent. His (receptive) ability in communicative situations therefore has to contain a good deal of linguistic competence. On the other hand his use of his communicative competence, when he has to understand a native speaker's words, can make up for deficiencies in his linguistic ability. So useful information may come from extra-linguistic factors such as the type of language-contact, from the setting, roles, topics, speech and/or listening intention. Therefore it would be artificial to separate the linguistic competence from the extra-linguistic factors that determine the communicative situation.

In the tests that are discussed here and that assess, only the listening or comprehension part of the communicative act, the testees are asked to demonstrate mastery of their linguistic competence. They are helped by the extra-linguistic guidance that comes from the specification of their role, of the topic, of the setting in that particular communicative act. And it is this guidance that we will consider to be the communicative context. It is meant to create or to substitute an authentic and intrinsical communicative intention, be it a speech intention or a listening one. Which Notions and Functions are used, is - partly - a consequence of the communicative intention.

4 How to assess listening comprehension

The communicative ability is often considered as the ability to convey information in the first place. The ability to receive information is given less attention. There are, of course, specific and exclusive listening situations. But apart from these, the importance of the receptive side in the communicative act is often underestimated. It is not the unpredictibility of the native speaker's reactions that tells strongly for the teaching and testing of listening comprehension. There are other arguments as well. The target-population of the tests we are discussing here consists of students in the initial phases of learning the language. Even if the most important objective of language learning was oral production as the main factor of communicative competence, there would be some arguments for training the receptive skills in the first years of the learning process in order to prepare in the most successful way the path towards oral proficiency. The so-called receptive semantic approach to language learning in the initial phase is gaining popularity. (Nord 1980, Postovsky in Winitz 1981)

However, listening and speaking are indeed very often integrated into a more complete communicative act. So we tried to take this fact into account by creating a communicative context and more particularly by guiding the specific listening intention as we explained before. For the tests this resulted in the simulation of (mini)-dialogues. The stimuli are supposed to be of such a low degree of difficulty that the testees could have produced them by themselves. The communicative need has to be immediately evident to them. The part to be tested is situated in the responses of the dialogues, the contents of which can be partly anticipated by the (easy) stimulus. The test-items focus on Notions or Functions.

5 The tests

Two types of tests will be described and demonstrated one after another.

5.1 The objective-based tests (General Notions)

For these tests 11 objectives were selected and formulated in terms of General Notions to be found in Un Niveau-Seuil (Coste 1976):

1 Interrogatives
2 Spatial Notions (location, relative position, distance, direction)
3 Temporal Notions (period, point of time)
4 Quantitative Notions (number, quantity)
5 Notions of degree
6 Notions of frequency
7 Deictic Notions (personal pronouns)
8 Deictic Notions (demonstrative pronouns/adjectives and possessive pronouns/adjectives
9 Deictic Notions (relative pronouns)

10 Contrastive Relations ((in)-equality, correspondence, contrast, comparison)
11 Logical Relations (con- and disjunction, in- and exclusion, cause, purpose, condition)

The overall goal of the tests is a formative one: to provide teachers and students (age-group 12-14 years) with a series of tests that can be used during the learning process to evaluate and to improve teaching-methods and learning-effects.

The learning-objectives are stated in behaviourial terms. General Notions from Un Niveau-Seuil were selected for three reasons. They occur in the curricula most frequently used for the beginning learners. They constitute a finite domain and they can reconciliate two opposite approaches: the discrete point and the communicative.

The language-forms belonging to a particular learning objective are, each, part of a test-unit. Each unit consists of a life-like situation or topic relevant to the target-group and is presented as a mini-dialogue. The stimulus - a question or a statement - is simple and is forced, as it were, upon the student as his own communicative need. An artificial procedure, it is true, but needed as a kind of advanced organizer of his listening intention.

The test-length varies from 5 to 20 units and the item-format that is used is multiple choice with 2 or 3 options preceded by a question, in the mother-tongue. Mastery of an objective is supposed to be reached at a test-score of about 80 per cent.

Objectives and their operationalisations into mini-dialogues and test-items have been pretested, revised and approved of in school-practice.

5.1.2 Some samples of objective-based tests

Learning objective 4: The testee must, in a short conversation, be able to understand notions that denote quantity.

Text on tape:	Test:
- Hé Michèle, combien d'argent as-tu dans ton portemonnaie?	How much money does Michèle's purse contain?
• Hélas, je n'ai pas d'argent du tout.	A There is no money in it at all. B There is still a lot in it.
- Les amis d'Henri étaient là, hier, au match de football?	Did Henry's friends go and watch the footballmatch?
• Il n'y avait que Paul. Les autres ne sont pas venus.	A No, only Paul did. B Yes, all of them did, except Paul.

Learning objective 10: The testee must, in a short conversation, be able to understand notions that denote comparison.

Text on tape:	Test:
- Tu as vu à la télé le match de tennis entre la France et l'Angleterre? ● Non, c'était bien? Est-ce que les Français ont perdu? - Oui, mais ils ont joué aussi bien que les Anglais.	What does the boy say about the Frenchmen's play? A They played better than the English. B They played just as well as the English. C They did not play as well as the English.
- Tu as beaucoup d'heures de français, cette année. ● Trois heures seulement. Ça fait une heure de moins que l'année dernière.	What does the girl say about the number of lessons in French this year? A There is one lesson per week more than last year. B There is one lesson per week less than last year. C There are just as many lessons per week as last year.

The underlined words are the language-forms belonging to the particular learning objective that is being assessed here.

5.1.3 Discrete point versus context

A major problem when constructing these tests is the question of context: does the context or the specific trait to be measured (time, comparison etc.) account for differences in scores? An answer to this question may be obtained from a comparison between each test and its parallel-test. For each learning objective two different tests have been constructed: test and parallel-test. Test and parallel-test contain the same elements of the learning objective, but differ in context and topic. General construction procedures have been the same.

Table 1 shows the scores on the test and parallel-test (pre-test version) dealing with the objective: logical relations. It is rather a long test and the figures may be considered representative of the test battery as a whole.

Table 1

Scores on the General Notions-test: logical relations, and on its parallel test

itemno.	$\bar{p}$	parallel $\bar{p}$	itemno.	$\bar{p}$	parallel $\bar{p}$
1	.97	.95	11	.86	.89
2	.91	.83	12*	.73	.96
3	.88	.88	13	.66	.78
4	.95	.81	14*	.92	.72
5	.88	.91	15*	.92	.51
6	.37	.38	16	.92	.96
7*	.90	.60	17	.97	.93
8	.84	.85	18	.86	.82
9	.84	.86	19	.96	.89
10	.88	.80			

Given the small differences in test scores between parallel-items one may conclude that in general, the global understanding of the contents is not such a strong factor that it outweighs the discrete point to be measured. In the cases preceded in the table by * disparallellism can be retraced to a lack or a surplus of redundancy in the context. In most cases redundancy has been reduced.

A confirmation of the non-global character of this way of testing may be found in the correlations between the separate tests each assessing different learning objectives, fluctuating from .27 to .55. These correlations show rather great differences. They are, as expected, generally lower than those between the subtests of the usual listening comprehension tests of Cito, that all aim at the assessment of global listening comprehension (r = circa .60).

5.2 Summative tests (Specific Notions and Functions)

We will now have a look at the second type of tests, the summative ones, that have been developed as subtests in a test battery for the lowest proficiency-level of Dutch secondary education. The whole of the battery was designed to assess the achievement of individual students after some 4 years of French.

5.2.1 Some history

The overall objective for listening comprehension-assessment at proficiency-level has been formulated as follows:

> "the ability to understand the foreign language, spontaneously spoken by educated native speakers at normal, conversational

speed and excluding extremely informal elements, lexical and syntactic elements that are incomprehensible to less educated native speakers and also excluding topics requiring specialistic knowledge". (Groot 1975)

For the assessment of this ability we have since the early 70s used one type of text: the interview subdivided into smaller parts, and one item format: the multiple choice-question consisting of one question and 3 answers, each presenting a one-phrase-summary of the global contents of each part. The items are formulated in the target-language.

Problems arose with respect to the finding of spontaneously spoken language (interviews) at the level under consideration. Other problems are connected with the exclusive implemention of the overall objective by interviews. There are far more listening situations and tasks than only those in interviews; and one can think of other item-formats as well. Validation research undertaken at our institute showed the constraints that the exclusive use of our multiple choice-technique imposed on the language material. (Van den Nieuwenhof 1979, De Jong 1982a/b) Diversification was to be found in both listening tasks/situations and question-types in order to enhance the validity of our tests. We tried to attain a greater variety on the basis of Un Niveau-Seuil, since we are dealing with the lowest level in our educational system. After having listed the possible listening situations/tasks, the topics and the item-formats, we made an enquiry among teachers about their and their students' preferences. These turned out to be the (traditional) interview, the dialogue or discussion, real life-like situations and samples of radio broadcast. As to item-format there was no real consensus.

5.2.2 Characteristics of the new tests

Three new tests such as students and teachers had shown preference for, were developed and tried out:

Information or messages passed on through the existing channels such as radio-broadcasts, preserving the characteristics of tone and language, and on topics suitable for the target-group.
The test requires testees to grasp the global contents of the topic (open listening intention) on the one hand and to understand detailed information like time, date, prices etc. (guided listening intention) on the other. The item-format is the multiple choice-question: 3 statements in the mother-tongue.

An every-day like situation in which the roles of the partners in the conversation are formal ones.
The test measures mainly mastering of Specific Notions (in the example given here, those having to do with buying clothes in a shop) as they are current with respect to a particular topic. The listening intention is therefore strongly guided. The item-format is multiple choice: 2 photographs, one of which represents the interlocutor's words.

A discussion that is characterized by the informal relations

between the partners. The topic is, again, supposed to be relevant to the target-group. The test is meant to assess the mastery of the Language-Functions that express emotions. The listening intention is guided towards the emotional or attitudinal aspects of the words spoken by each of the partners. The student is, as it were, overhearing a conversation, between youngsters of the same age-groupe. The item-format is multiple choice: 3 statements in the mother-tongue.

These new tests have been integrated in a test battery which also includes a traditional interview.

5.2.3 Some examples from the every-day like situation test: shopping (Specific Notions)

We have selected an example from this particular test because it corresponds most with the discrete point approach in a communicative context. The listening intention is strongly guided. Testees are asked to identify themselves with the client-part and to focus, as far as the test is concerned, on the answers of the shop assistant. These become clear in one out of two photographs.

La cliente: - Le pantalon bleu dans la vitrine, je peux le voir s.v.p.?
La vendeuse: • Mais bien sûr; venez avec moi, si vous voulez; vous allez me montrer lequel, car il y a deux pantalons bleus dans la vitrine.

A

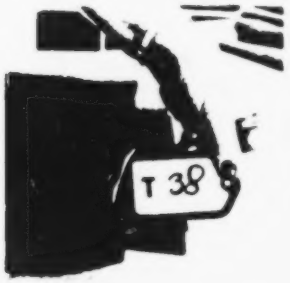
B

La cliente: - Je peux essayer ce pantalon en taille 40, s.v.p.?
La vendeuse: • Bien sûr. Les cabines d'essayage sont derrière vous, regardez ... là-bas.

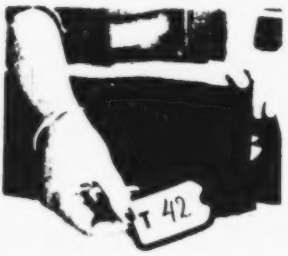
A

B

La cliente: - Ça ne va pas.
La vendeuse: • Ce modèle taille très petit. Voilà, essayez la taille au-dessus. Le 42 vous ira certainement mieux.



A



B

5.2.4 Problems

Photographs are used in this test for various reasons. We wanted to diversify our item-format staying however within the range of the closed question-types. Photographs have the advantage of avoiding language-problems like having to read statements either in the target-language or in the mother-tongue. At the same time they give some of the extra-linguistic support a non-native speaker would have made use of in the real situation.

In table 2 the data of this test are given in relation to those of the other tests of the battery in order to enable us to make a comparison. The test we are discussing here is half of the test that is labelled A. (A2)

Table 2

Results on the Specific Notions-test compared to the rest of the test battery

	n	$\bar{x}$	$\bar{p}$	SD	r (KR20) tt
A1* A2	23(23)	19.35(17.85)	.84(.78)	2.17(2.13)	.46(.25)
B*	12(12)	10.14(9.20)	.85(.77)	1.66(1.63)	.50(.36)
C*	15(15)	12.57(10.60)	.84(.77)	1.98(1.[illegible])	.53(.38)
A B C	50(50)	42.06(38.67)	.84(.77)	4.41(4.33)	.70(.60)

Note: N = 972(107)
* = the separate parts of the test (the subtests)
A1 = radio information, A2 = Specific Notions (shopping)
B = interview, C = discussion containing emotional and attitudinal Functions

Test A proved to be less reliable than our usual tests of which B is a representative specimen. Test C shows more or less the outcome of our standard tests which is not very surprising since test C is in its form closer to the normal interview-type.

Reliability, however, is a problematic concept in the outcome of this specific test battery. Although test A, B and C are supposed to measure testees' listening comprehension at the lowest level of the Dutch secondary educational system, this battery is always taken together with a set of tests meant for one level higher. Students having followed the same curriculum can take examinations on two different levels. As to listening comprehension, the decision procedures to assign students to one level or to the other, are based on the assessment of all students' listening comprehension on both levels. This means that a large proportion of students belonging to the next higher level distort the data shown in table 2. There is no other means of distinguishing in advance between the two levels in the population. Finally about 20 per cent of the population in table 2 is to be found on the lower level we aim at in the tests discussed here. In table 2 the figures of the actual target-populations for which these tests have been developed, are given in brackets.

The fact that reliability is even lower here than for the whole population is not surprising since the normal distribution cannot be expected. After all, students belonging to this group (in brackets) are the lowest scoring testees of the whole population of table 2. Nevertheless the test battery developed for the lowest level in combination with the tests for the higher level comply with the objective to make a distinction between the two levels with a sufficient degree of certainty ($p < .01$) (see Appendix)

Yet test A remains a problem, especially the photo-section. It hardly contributes to the assessment of testees' listening comprehension. The item-format might account for weakness. At this moment further validation research is undertaken on this particular issue. The intrinsic reliability of this test will be measured by a comparison with two other item-format for the same contents (text).

6 Conclusion

We have discussed and shown examples of two different types of listening comprehension tests based on different Notions of the T-Level and Un Niveau-Seuil.

The first type, learning objective-based tests, has been developed to enable teachers and students, in the initial phase of the learning period, to find out about mastery of discrete points (General Notions) in a meaningful context. We think that we have succeeded in reconciliating, to a certain extent, two seemingly opposite approaches.

As to the second type, summative listening tests we have shown our efforts to integrate an other kind of discrete-point testing (Specific Notions) into a broader communicative context, together with a variety in item-formats of the closed question-type. As far as the photo-test is concerned, we encountered problems of reliability, but we think this kind of tests might be an enrichment of our test battery as regards both contents and item-format. Further development of this technique together with further validation research seems worthwile.

Appendix (with acknowledgement to my colleague John de Jong)

For this particular school type final examinations are administered at two different levels. Most pupils aim at the higher level and in fact 80 per cent succeed in taking this higher level. However, decisions on the level at which pupils graduate , are based on their achievements on tests they have already taken. This is why all pupils of this school type partake in tests of both levels. Depending on their results the final decision on their graduation level is taken and consequently only the results of those tests that are meant to assess that particular level are taken into account. In this situation it is essential that tests for each of the levels should clearly differ in required ability. Data on classical test analysis of both tests are presented in table 3.

Table 3

Results of one population on two tests of different levels

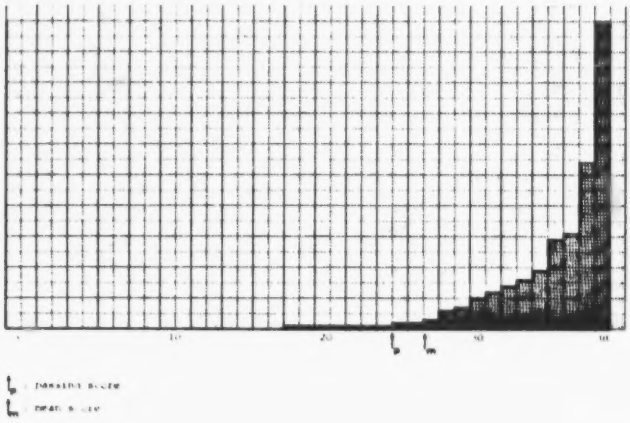
	higher level tests	lower level tests
$\underline{n}$	38	50
$\underline{p}$	.72	.84
SD	4.41	4.41
r (KR20) tt	.69	.70

Note: N = 972

Product-moment-correlation between the two tests is .73. Classical test analysis including p-m-correlation, however, gives a global evaluation of test quality only and does not take into account discriminating power at different sub-levels and consequently neglects differences in standard error at different ability levels. Of course, one would expect higher scores for alle pupils on the easier (i.e. lower level) test. But pupils who succeed on all items in the higher level test cannot be expected to obtain an even higher proportion of right answers on the lower level test. To illustrate this point, figure 1 presents - at each different score in the higher level test - violations against the expectation of a higher score on the lower level test. Violations are defined here as pupils obtaining a higher or equal proportion of right answers on the higher level test. The violations are expressed as a proportion of the frequency distribution: the total number of pupils obtaining that particular score. In the critical area below the passing score less than .01 violation can be observed. It is only well above the mean score that violations constitute a considerable proportion of the observed scores and all pupils with a maximum score on the higher level test obtained an equal high score on the easier test. Misclassifications of pupils appear to occur less than indices from classical test analysis would suggest.

Figure 1

Violations - as proportion of observed scores - (y-axis) versus score on higher level test (x-axis)



Bibliography

Canale, M., Swain, M. 1980. "Theoretical basis of communicative approaches to second language teaching and testing." Applied Linguistics 1: 1-47.

Coste, D. et al. 1976. Un niveaul-Seuil. Strasbourg: Council of Europe.

Ek, J.A. van, 1976. The threshold level for modern language learning in schools. Strasbourg: Council of Europe.

Groot, P.J.M. 1975. "Testing communicative competence in listening comprehension." In Jones, R. and Spolsky, B. (eds.) 1975. Testing language proficiency. Arlington, Virginia: Center for Applied Linguistics.

Jong, J.H.A.L. de, 1982a. Testing foreign language listening comprehension. Paper presented at ISET 5, Stirling Scotland.

Jong, J.H.A.L. de, Nieuwenhof, H.W.M. van den, 1982b. Een experimentele luistervaardigheidstoets. Arnhem: Cito.

Nieuwenhof, H.W.M. van den, Jong, J.H.A.L. de, et al. 1979. Validatie van luistertoetsen moderne vreemde talen. Arnhem: Cito.

Nord, J.R. 1980. "Developing listening fluency before speaking: an alternative paradigm." System 8: 1-22.

Postovsky, V.A. 1981. "The priority of aural comprehension in the language acquisition process." In Winitz, H. (ed.) The comprehension approach.

Winitz, H. (ed.) 1981. The comprehension approach to foreign language instruction. Rowley, Massachussets: Newbury House.